

UNITED STATES DISTRICT COURT
DISTRICT OF MASSACHUSETTS

SCANSOFT, INC.,)	
)	
)	
Plaintiff,)	
)	
v.)	C.A. No. 04-10353-PBS
)	
)	
VOICE SIGNAL TECHNOLOGIES, INC.,)	
LAURENCE S. GILICK, ROBERT S.)	
ROTH, JONATHAN P. YAMRON, and)	
MANFRED G. GRABHERR,)	
)	
Defendants.)	

**VOICE SIGNAL TECHNOLOGIES, INC.'S OPENING CLAIM
CONSTRUCTION MEMORANDUM FOR U.S. PATENT 6,594,630**

Defendant and counterclaim plaintiff Voice Signal Technologies, Inc. ("Voice Signal") submits this memorandum in support of its proposed construction of the claims of United States Patent Number 6,594,630 (the '630 patent).

In accordance with the Order of the Court issued at the March 16, 2005 hearing, the parties have identified the claim terms that should be construed and have exchanged proposed constructions. Voice Signal will demonstrate that, under controlling law and in view of the intrinsic claim construction evidence, the interpretations advanced by Voice Signal should be adopted.

I. STANDARD FOR CLAIM CONSTRUCTION

Claim construction is an issue of law that is resolved by the Court. *Markman v. Westview Instruments, Inc.*, 52 F.3d 967, 976 (Fed. Cir. 1995). When construing a patent claim, courts

must look first to the three sources of intrinsic evidence – the patent’s claims, written description and prosecution history. *Genentech v. Boehringer Mannheim*, 989 F. Supp. 359, 363 (D. Mass. 1997). *See also Vitronics Corp. v. Conceptronic, Inc.*, 90 F.3d 1576, 1581-82 (Fed. Cir. 1996).

The starting point in claim construction is the language of the claim itself, which sets forth the scope of the claimed invention unless the written description requires a contrary reading. *VLT Corp. v. Lambda Electronics, Inc.*, 283 F. Supp.2d 347, 349 (D. Mass. 2003), citing *Northern Telecom Ltd. v. Samsung Electronics*, 215 F.3d 1281, 1287 (Fed. Cir. 2001). Terms in a claim should be given their ordinary or customary meaning. *Vitronics*, 90 F.3d at 1582. The ordinary meaning of a claim must be determined “from the standpoint of a person of ordinary skill in the relevant art.” *Harvest Tech. Corp. v. Cytomedix, Inc.*, 2004 WL 2009253, *3 (D. Mass. 2004) (citing *Teleflex, Inc. v. Ficosa N. Am. Corp.*, 299 F.3d 1313, 1325 (Fed. Cir. 2002)).

Claims must be read in the context of the specification of which they are a part. *Vitronics*, at 1582. As this court has noted, “because many terms in a claim involving a complex patent are construed from the perspective of one skilled in the art at a point of time in the past, and because most judges don’t have those qualifications, judges will often have to turn to the specification to understand the claim terms.” *VLT*, 238 F.Supp. 2d at 351. The specification is widely recognized as “the single best guide to the meaning of a disputed term.” *Genentech*, 989 F. Supp. at 363, *Vitronics*, 90 F. 3d at 1582.

Similarly, the prosecution history, as the “undisputed public record of proceedings in the Patent and Trademark Office, is of primary significance in understanding the claims.” *Vitronics*, 90 F. 3d at 980 (internal quotation and citations omitted). In particular, representations made by the patentee to the Patent and Trademark Office regarding the scope of the claim control the

interpretation of the claim language in the issued patent. *Southwell Tech., Inc. v. Cardinal IG Co.*, 54 F.3d 1570, 1576 (Fed. Cir. 1995).

Where the intrinsic evidence of claim construction is insufficient to interpret a claim term or limitation, the Court may resort to extrinsic evidence, including dictionaries, learned treatises and expert testimony. *Dow Chem. Co. v. Sumitomo Chem. Co., Ltd.*, 257 F.3d 1364, 1373 (Fed. Cir. 2001). Both this Court and the Federal Circuit have acknowledged, “dictionaries, encyclopedias and treatises are particularly useful resources to assist the court in determining the ordinary and customary meanings of claim terms.” *Harvest Tech.* 2004 WL 2009253, *4 (citing *Texas Digital Sys., Inc. v. Telegenix, Inc.*, 308 F.3d 1193, 1204 (Fed. Cir. 2002)). In addition, the court may rely on expert testimony to understand the technology at issue and the ordinary meaning of terms to practitioners in the art. *VLT*, 238 F. Supp.2d at 350, *Vitronics*, 90 F. 3d at 1583-84.

Finally, claims should be construed to uphold their validity, if possible. *Modine Mfg. Co. v. U.S. Int’l Trade Com’n*, 75 F.3d 1545, 1557 (Fed. Cir. 1996) (rejecting indefiniteness argument after construing claims, stating that “when claims are amenable to more than one construction, they should when reasonably possible be interpreted so as to preserve their validity”). *See also Whittaker Corp. v. UNR Indus. Inc.*, 911 F. 2d 709, 712 (Fed. Cir. 1990).

II. BACKGROUND – SPEECH RECOGNITION AND THE '630 PATENT

In general terms, the invention of the '630 patent is an apparatus and method for allowing a user to control an electrical device using voice commands. An example used in the written description is the command “lights on.” The invention involves recognizing the command,

controlling the electrical device in response to the recognized command, and preventing operation of the device if the command is not sufficiently distinct from other noise.

There are many known ways to perform speech recognition. All of them involve matching a received utterance against a library of pre-recorded utterances (referred to in the patent as “models” or “templates”) to identify the closest match. Thus, in the simple example given in the patent, the command “lights on” is “recognized” by comparing it to the pre-stored templates “lights on” “lights off” and “dim lights” to calculate a score for the closest match, which should, of course, cause the system to recognize the utterances “lights on.”

One of the challenges in speech recognition, particularly in a noisy environment, is “wordspotting.” The '630 patent explains that “the recognizer must spot a keyword embedded in other speech or sounds...while at the same time reject speech that does not include any of the valid keywords,” including background noise. '630 patent, col. 1, l. 41-44. The patent also states that the greater the number of syllables contained in a keyword, the easier it is to spot the keyword. '630 patent, col. 1, l. 47-53. In the process of matching the received sounds against the pre-recorded templates, the system has more information to match (and thus will be more accurate) as the number of syllables increases.

However, as the patent explains, “while using longer voice commands provides an easy way of boosting the performance of wordspotters, it is more convenient for users to memorize and say short commands.” '630 patent, col. 1, l. 54-57. Therefore, while the wordspotting would be more accurate if the command for turning the lights on were “please turn the lights on now,” users are more likely to remember and use a shorter command, such as “lights on.”

The invention of the '630 patent combines increased accuracy and user convenience by using the pauses, which occur naturally between and around simple command words, to increase

the data available for matching. Thus, for example, the command “lights on,” which we recognize as two one-syllable words or a two-syllable phrase, is recognized by the inventive system as a three-syllable phrase: “lights <pause> on.” Three-syllable phrases are more reliably matched than two-syllable phrases. '630 patent, col. 3, l. 59-61. Similarly, the '630 patent teaches capturing the natural pauses that occur before and after the command words, thus extending the two-syllable phrase, “lights on,” to a five-syllable phrase: <pause> lights <pause> on <pause> '630 patent, col. 10, l. 15-18.

The '630 patent also teaches comparing command words and pause portions of an audio command to other noise, or background noise. This allows the inventive method to prevent an electrical device from being activated if it has not clearly received an audio command (as opposed to other noises or background noise). It also permits the system to recognize audio commands, even where the background noise is “loud.” Thus, the user can enter audio commands in a loud environment, so long as the command words are louder than the background noise. '630 patent, col. 12, l. 4-7.

Voice Signal has asserted two independent claims (cl. 7 and 16) and has reserved its rights to assert one or more of the claims which depend from claim 7.¹ ScanSoft has identified all of claims 7, 14 and 16 as in need of construction. Voice Signal addresses each term below.

III. CLAIM CONSTRUCTION

Claim 7

Claim 7 reads, in its entirety, as follows:

¹ Voice Signal cannot make infringement determinations until its expert receives and reviews ScanSoft's source code. ScanSoft has acknowledged that it must produce the source code for its accused products, but has steadfastly refused to produce its source code on the untenable ground that it should not have to produce it unless and until Voice Signal produces its source code. ScanSoft has not sought a protective order. Voice Signal has obtained an order compelling production and has filed a motion for sanctions. Once it receives and reviews ScanSoft's source code, Voice Signal will identify the dependent claims asserted.

A method of activating an electrical device through at least one audio command from a user, the method comprising the steps of:

- [1] recording speech recognition data having a command word portion and a pause portion, each of the speech recognition data portions being at least one syllable in length;
- [2] receiving at least one audio command from a user, the at least one audio command having a command word portion and a pause portion, each of the audio command portions being at least one syllable in length;
- [3] comparing said command word portion and said pause portion of said at least one received audio command with said command word portion and said pause portion, respectively, of said speech recognition data;
- [4] generating at least one control signal based on said comparison;
- [5] controlling power delivered to an electrical device in response to said at least one control signal for operating the electrical device in response to said at least one received audio command;
- [6] analyzing the pause portion of the received audio command for spectral content; and
- [7] preventing operation of the electrical device when the spectral content is dynamic.

Claim Term	ScanSoft's Proposed Construction	Voice Signal's Proposed Construction
A method of activating an electronic device through at least one audio command from a user, the method comprising the steps of	<p>A process for turning on an electrical device by an audio command.</p> <p>An "audio command" is a word, phrase, or numeric digit, alone or in combination with other words, phrases, and/or digits, which is spoken by a user to turn on the electrical device.</p>	<p>"Activating" means "to make active or more active." Therefore, a process for making an electronic device active or more active.</p> <p>"audio command" means an audible combination of command words and pauses that activate an electrical device.</p>

“Activate” is a common word with a plain meaning. It is not limited to turning something on. In the art, the word activate means to engage the control process of a device, whatever that control mechanism is. Declaration of Charles C. Wooters (“Wooters Aff.”), ¶ 3. It includes turning something on or acting on something that is already on to make it more active, less active, or turn it off.

“Audio” means “of or relating to acoustic, mechanical, or electrical frequencies corresponding to the normally audible sound waves,” or “audible sound.” *Webster’s Third New International Dictionary (1993)*. The claim itself defines an “audio command” to include a command word portion and a pause portion. Thus this part of the claim requires a method of controlling an electrical device by the use of speech that has both a command word portion and a pause portion.

Contrary to ScanSoft’s interpretation, the audio command is not limited to a command to turn on an electrical device. In fact, the specification teaches that the audio command might be used to turn on a device, adjust the power to a device upward or downward, or turn off a device. *See* '630 patent, col. 7, l. 26-33.

Claim Term	ScanSoft’s Proposed Construction	Voice Signal’s Proposed Construction
[1] recording speech recognition data having a command word portion and pause portion, each of the speech recognition data portions being at least one syllable in length;	<p>The “speech recognition data” is the speech model information that unknown input speech is compared to.</p> <p>A “pause” refers to a momentary stop in an ongoing event. The patent and its prosecution history define a syllable as a length of time at least about 200 milliseconds long. A “pause portion” is that part of speech recognition data comprising a pause.</p>	<p>“speech recognition data” is the speech model data, or templates, to which a received audio command is to be compared.</p> <p>the “pause portion” of the speech recognition data is the speech recognition data, or templates, for the portion of the audio command that is not the command word portion.</p> <p>“at least one syllable in</p>

Claim Term	ScanSoft's Proposed Construction	Voice Signal's Proposed Construction
	Thus, this claim element requires that the speech model information includes a momentary stop in speech that is at least about 200 milliseconds long.	length" means "within the duration range for one syllable of speech."

"Speech Recognition Data"

"Speech recognition data" is the pre-stored template that models the sounds of human speech and other sounds (such as background noise). The speech recognition data is stored in the speech recognizer. The received audio commands are compared to the speech recognition data. Thus, "speech recognition data," should be construed to mean "the speech model data to which the received audio command is to be compared." The speech recognition data (like the audio command) has command word portions (*e.g.*, "lights" and "on") and pause portions.

ScanSoft's proposed construction of "speech recognition data" introduces confusion, not clarity. It introduces the terms "unknown input speech," rather than using "received audio command," the term found in the claims.

"At Least One Syllable In Length"

"At least one syllable in length" means at least the length of a spoken syllable. Speech scientists have studied the length of spoken syllables in the English language and in other languages. Because of accents and idiosyncratic speech habits, the duration of a syllable is different from speaker to speaker. For example, the one-syllable word "run" might be pronounced "run" (*e.g.*, by a northerner) or "ruuuuun" (*e.g.*, by a southerner). *See* '630 patent, col. 3, l. 9-11. The duration of a syllable also varies for any given speaker, even when saying the same combination of words. Wooters Decl. ¶ 4. By studying statistically significant pools of

speakers, scientists have arrived at an average syllable length of about 200 milliseconds, with a standard deviation of 100 milliseconds. *Id.* ¶ 4. In other words, most syllables are between 100 and 300 milliseconds in duration. *Id.* ¶ 4.

The claim limitation “at least about one syllable in length” is used throughout the '630 patent claims, with respect to both the command word portions and the pause portions of speech recognition data (the stored templates) and audio commands (the received speech designed to activate an electrical device). It should be interpreted to mean command word portions or pause portions, the duration of which falls within this known range of a syllable, i.e. “the duration of which is within the duration range for one syllable of speech.”

ScanSoft urges the Court to adopt the construction “about 200 milliseconds,” presumably relying on an illustration of the invention in the written description found at column 10, lines 5-8 and 23-24, and ignoring references at col. 3, l. 56-57, 65-66; col. 4, l. 27, 30, 47 and 51. The claim should not be so limited. To do so would be to read a limitation into the claim from the written description where there is no need or justification for doing so.²

ScanSoft’s interpretation would also be directly at odds with the prosecution history.³ In his initial office action on the application that became the '630 patent, the examiner objected to the disclosure, in part, because

[t]he disclosure that a *pause may be substituted for a syllable* will not support either a claim where *the pause is measured in terms of syllables*, nor a potential claim of a *pause consisting of multiple syllables*. A more appropriate unit to designate measurement and duration, or for the representation of multiple pauses should be used.

² In any event, the “about” in “about 200 milliseconds” proposed by ScanSoft would have to be the known range around the 200 milliseconds mean, as shown in the figure above, which is the range from 100 milliseconds to 300 milliseconds.

³ A full copy of the prosecution history of the '630 patent is submitted with the Declaration of Wendy Plotkin, filed herewith. References to particular portions of the prosecution history are by control numbers (e.g. PH001) placed on the file by Voice Signal’s counsel.

PH 65. The applicants “respectfully disagree[d]” with the examiner, noting that “[a] syllable is the perfect measure of a pause.” *Id.* at 87. The applicants pointed out that syllables may have quite different durations depending on the surrounding words and/or the particular speaking style of the user. *Id.* at 88-89. The applicant provided evidence that, in speech, (1) a syllable is a commonly accepted measure of the duration of a pause; and (2) in terms of measuring the length of a pause, the syllable is a clear alternative for using the units of time (such as minutes) for the same purpose. *Id.*

The examiner agreed with applicants and withdrew his objection, stating:

The material provided does indeed support that a *pause may be measured in syllables*, and the Examiner stands corrected.

Id. at 154.

“Pause Portion”

The words “pause portion” (in the speech recognition data) is the special recognition data, or template, for the portion of the audio command which is not the command word portion. ScanSoft’s proposed construction asks the Court to replace the common term “pause” with the words “momentary stop in speech.” There is no support in the written description or the prosecution history for this interpretation. Furthermore, it is contrary to the gist of the invention. The pause is simply the natural space between command words. It is, and the patent treats it as, a component of speech.

The prior art taught a method of “discrete dictation,” which required the speaker artificially to stop speaking between words and used those interruptions to identify the beginning point and endpoint of keywords in a “wordspotting” technique. *See* '630 patent, col. 1, 1. 64 – col. 2, 1. 19. The invention of the '630 patent is not a discrete dictation system. It does not require the user artificially to stop speaking in between each word. It does not use the data

between command words (the “pause portion”) to find the beginning and end points of the command words. The pause portion(s), as used in the invention of the '630 patent, are not “stops” in speech. Rather, the pause portion(s) are an integral part of the speech data collected and the audio command to be analyzed. *See* '630 patent, col. 3, 1. 62 - col. 4, 1. 9.

Therefore, the court should define the term “pause portion” in this limitation to mean the speech recognition data, or template, for the portion of an audio command that is not the command word portion. It should reject ScanSoft’s suggestion that “pause” means a “momentary stops in speech.”

Claim Term	ScanSoft’s Proposed Construction	Voice Signal’s Proposed Construction
[2] receiving at least one audio command from a user, the at least one audio command having a command word portion and a pause portion, each of the audio command portions being at least one syllable in length;	The system gets from a user an audio command that has two parts: (1) a word part, and (2) a pause part, each of which must be at least 200 milliseconds long.	<p>“audio command” means an audible combination of command words and pauses that activate an electrical device.</p> <p>“at least one syllable in length” means “within the duration range for one syllable of speech.”</p>

Voice Signal’s claim construction positions regarding “audio command” and “at least one syllable in length” are set forth at pages 7-11 above.

Claim Term	ScanSoft’s Proposed Construction	Voice Signal’s Proposed Construction
[3] comparing said command word portion and said pause portion of said at least one received audio command with said command word portion and said pause portion, respectively, of said speech recognition data;	An input command having a word portion of at least one syllable in length, ~200 milliseconds, and a pause portion of at least one syllable in length, ~200 milliseconds, is compared to speech recognition model data having a word portion at least one syllable in length, ~200 milliseconds, and a pause	Comparing the command word portion and the pause portion of the received audio command to the command word portions and the pause portions, respectively, of the speech recognition data.

Claim Term	ScanSoft's Proposed Construction	Voice Signal's Proposed Construction
	portion at least one syllable in length, ~200 milliseconds.	

ScanSoft urges the Court to “construe” this claim term by reading into the limitation a host of words not found in the claim.

The claim term means what it says: The system compares the received audio commands (which include command word portions and pause portions) to the previously stored speech recognition data for audio commands (which includes templates, or models, for command word portions and pause portions). It should be construed as such: “comparing the command word portion and the pause portion of the received audio command to the command word portions and the pause portions, respectively, of the speech recognition data.”

Claim Term	ScanSoft's Proposed Construction	Voice Signal's Proposed Construction
[4] generating at least one control signal based on said comparison;	A “control signal” is a signal that directs or manages the operation of the electrical device that is controlled by the control signal.	Voice Signal accepts ScanSoft's proposed construction.

Claim Term	ScanSoft's Proposed Construction	Voice Signal's Proposed Construction
[5] controlling power delivered to an electrical device in response to said at least one control signal for operating the electrical device in response to said at least one received audio command;	Electrical power is turned on or off, or increased or decreased, to an electrical device in response to a control signal generated in response to an input audio command.	Electrical power is turned on or off, or increased or decreased, to an electrical device in response to a control signal generated in response to a received audio command.

The parties' constructions are almost identical. Voice Signal's is preferable because it ties to the language of the claims, using the words “in response to a received audio command”

instead of “in response to an input audio command.” Input audio command is not a term used in the claims and should not be brought in through construction.

Claim Term	ScanSoft’s Proposed Construction	Voice Signal’s Proposed Construction
[6] analyzing the pause portion of the received audio command for spectral content; and	“Content” refers to what is contained within something. In this case, the content of the pause portion of the received audio command must be “spectral.” A speech signal contains spectral components in the sense of energy at various different specific frequencies. Thus, “spectral content” means that the pause portion of the received audio command must have energy at various different frequencies.	Sounds are made up of energy at different frequencies. Spectral content is the amount of energy in each frequency that makes up a sound or a series of sounds over a specified period of time.

Sound travels through the air in waves. Each wave has a frequency (measured as the distance between the repeating peaks of the waves) and an energy, or amplitude (which can be thought of as the height of the peaks of the waves). Any given sound consists of a combination of these waves. In very simple terms, one can think, for example, of a “high” sound (a note sung by a soprano) as having relatively more energy in the high frequencies, and relatively less energy in the low frequencies. The corollary is also true. A “low” sound (a note sung by an alto) has relatively more energy in the low frequencies and relatively less energy in the high frequencies. Wooters Decl. ¶ 5.

“Spectral content” is the amount of energy at each frequency that makes up a sound or series of sounds over a specified period of time. All sound waves have energy at different frequencies. Therefore, all sound (including command words, pauses, and background noise) has spectral content. Wooters Decl. ¶ 5.

Claim Term	ScanSoft's Proposed Construction	Voice Signal's Proposed Construction
[7] preventing operation of the electrical device when the spectral content is dynamic.	"Dynamic" refers to something that is in motion or is changing. Thus, this claim element means that operation of the electrical device is prevented when the at least one syllable long pause portion of the input audio command has energy at various different frequencies and which is moving or changing.	When the change in the spectral content of the pause portion is different from the change in spectral content that would be expected in the background noise, the spectral content is said to be dynamic. If the spectral content of the pause portion of the received audio command is dynamic, the electrical device is prevented from operating.

Both parties agree that "dynamic" means changing. In the real world, the spectral content (energy at different frequencies) is always changing, even in "quiet" environments. For example, even if no one is speaking in an office environment, the background noise includes air handling systems, fluorescent lights, etc. The spectral content of even that "quiet" background noise is always changing. Wooters Decl. ¶ 6. Therefore, spectral content is always "changing." In speech recognition, "dynamic" in this context means change in the spectral content that is different from what would be expected in the background noise. *Id.* ¶ 6.

Voice Signal's proposed construction is consistent with the written description of the '630 patent. The written description clearly describes the invention as analyzing the activity of the spectral content of the pause portion of the audio command in comparison to the spectral content of the background noise. It teaches that the change in the spectral content of the pause portion shall be compared to a preset level, so as to permit the system to operate if the change in the spectral content of the pause is within a pre-set range of the change in spectral content expected in the background noise. Correspondingly, it describes a pre-set range of change in spectral content above what would be expected in the background noise (i.e. the pre-set level expected for keywords). If the amount of change in the spectral content of the pause is in this pre-set

range above background noise, the system determines that a pause has not occurred and, therefore that it has not received an audio command (which must have at least one pause portion) and prevents operation of the electrical device. *See* '630 patent, col. 10, l. 43 - col. 11, l. 11.⁴

If the claim were interpreted without reference to the background noise, the claimed system would never activate an electronic device. *Wooters Deal* ¶ 7. It thus would defeat the stated purpose of the invention. ScanSoft proposes just this construction, suggesting that, if there is any change in spectral content during the pause portion of the received audio command, the device will be prevented from operating. This is not only inconsistent with the written description, but would also render the claimed invention useless. Claims should not be construed to render the purpose of the invention impossible. *See Chimie v. PPG Indus., Inc.*, 402 F. 3d 1371, 1377 (Fed. Cir. 2005) (refusing to construe term “dust free and non-dusting” in overly literal manner that would exclude the claimed device); *Carroll Touch, Inc. v. Electro Medical Systems, Inc.*, 15 F. 3d 1573, 1578 (Fed. Cir. 1993). *See also Bausch & Lomb, Inc. v. Barnes-Hind/Hydrocurve, Inc.*, 796 F.2d 443, 450 (Fed. Cir. 1986) (rejecting definition of “smooth” as “absolutely ridge free”); see also *Lucas Aerospace, Ltd. v. Unison Ind., L.P.*, 890 F. Supp. 329, 335 (D. Del. 1995) (holding, as a matter of law, that court cannot accord a meaning to a term that is physically impossible to achieve).

Claim 14

Claim 14 reads, in its entirety, as follows:

The method of claim 7, and further comprising:

⁴ This interpretation is confirmed by dependent claims 9 and 14, each of which include as a limitation “receiving background noise data in conjunction with receiving [] at least one audio command.” It is axiomatic that a dependant claim must be subsumed within the independent claim on which it depends. *See Wolverine World Wide, Inc. v. Nike, Inc.*, 38 F. 3d 1192, 1199 (Fed. Cir. 1994). Therefore, it is clear that the method of claim 7 includes receiving background noise data in conjunction with receiving [] at least one audio command.

- [1] receiving background noise data in conjunction with receiving said at least one audio command;
- [2] ascertaining a first energy content for the command word portion of the received audio command;
- [3] ascertaining a second energy contents for the received background noise data;
- [4] comparing the first and second energy contents and generating a corresponding energy comparison value; and
- [5] preventing the generation of said at least one control signal when said energy comparison value is below a predetermined level.

Claim Term	ScanSoft's Proposed Construction	Voice Signal's Proposed Construction
[1] receiving background noise data in conjunction with receiving said at least one audio command;	"Background noise" is part of the user's acoustic environment that includes everything besides the spoken user input. Thus, an acoustic signal input is received that includes a spoken command from the user, and everything else in the background acoustic environment.	"Background noise" is all sound which is not the audio command. It may include voice conversations, machinery, television, radio, and the like, or any combination thereof.

Background noise is described in the specification. It is the noise existing in the background, which is to be distinguished from the commands. The written description states that background noise may include noise "such as voice conversations, machinery, television, radio and the like or any combination thereof." '630 patent, col. 8, l. 21-22; col. 9, l. 35-37, 58-60.

Inherent in the task of recognizing audio commands is the difficulty of distinguishing audio commands from background noise, including conversations or other loud noises present in the environment. The system described and claimed in the '630 patent is essentially an open microphone, which constantly receives background noise and occasionally receives audio

commands. *See* '630 patent, col. 8, l. 19-25. The claims of the '630 patent distinguish between an audio command and background noise.

ScanSoft's proposed construction reads in new words to the claim not found in the specification or anywhere else. It introduces the terms "acoustic environment" and "acoustic signal input" which are not defined and do not have an obvious meaning. They are not found in the specification. The Court should adopt Voice Signal's construction, which is rooted in the specification.

Claim Term	ScanSoft's Proposed Construction	Voice Signal's Proposed Construction
[2] ascertaining a first energy content for the command word portion of the received audio command;	<p>"Content" refers to what is contained within something. In this case, the content of the command word portion of the received audio command must be 'energy' content.</p> <p>The energy of a speech signal refers to its force or strength and in visual representations is characterized by the amplitude or size of the speech signal.</p> <p>Thus, "first energy content" means that the command word portion of the received audio command is characterized by a first kind of inherent force or strength within it.</p>	<p>"energy content" is the total amount of energy detected in a given segment of sound.</p> <p>"first energy content" is the energy content of the command word portion of the received audio command.</p>

As explained in the written description, one mechanism used by the claimed invention to distinguish audio commands from background noise is to compare the total energy of the command word portion of the audio command to the total energy of the background noise. Energy content generally correlates to what we think of as loudness. Wooters' Decl. ¶ 8. The claimed method compares the energy of the command word portion of an audio command to the

energy content of the background noise. If the difference between the energy content of a command word and the energy content of background data is below a pre-established threshold, no control signal may be given (e.g. the lights are not turned on). '630 patent, col. 11, l. 50 – col. 12, l. 3. As the written description explains, “[w]ith the above-described arrangement, the user can enter commands even in loud environments by talking louder than the background.” '630 patent, col. 12, l. 4-6.

The first step in this process is to ascertain the first thing to be compared - - the energy content of the command word portion of the received audio command. This is called the “first energy content” in claim 14.

ScanSoft’s proposed construction introduces the terms “inherent force or strength” into the claim. This is vague, ambiguous and unnecessary. Energy is a known and measurable variable. It generally corresponds to loudness. In the science of speech recognition, it is typically measured in decibels. Wooters Decl. ¶ 8.

Claim Term	ScanSoft’s Proposed Construction	Voice Signal’s Proposed Construction
[3] ascertaining a second energy content for the received background noise data;	The system must also determine a “second energy content,” which is characteristic of the inherent force or strength contained within the background noise signal	<p>“energy content” is the total amount of energy detected in a given segment of sound.</p> <p>“Background noise data” is the background noise, which may include noise such as voice conversations, machinery, television, radio, and the like, or any combination thereof.</p> <p>“Second energy content” is the energy content of the received background noise data.</p>

The second step in comparing the energy, or loudness, of the audio command to the background noise, is to ascertain the energy content of the background noise. As explained above, the background noise is the noise which is not the audio command. It may include noise such as “voice conversations, machinery, television, radio and the like, or any combination thereof.” '630 patent, col. 8, l. 21-22. The measure of the energy content of the background noise in claim 14 is the “second energy content.”

ScanSoft again inserts “inherent force or strength” into the claim. Again, these terms are vague, ambiguous and unnecessary.

Claim Term	ScanSoft's Proposed Construction	Voice Signal's Proposed Construction
[4] comparing the first and second energy contents and generating a corresponding energy comparison value; and	A signal is generated that has a characteristic value representative of the correspondence between the first and second energy contents	comparing the first energy content (of the command word portion of the received audio command) with the second energy content (of the background noise data) and arriving at an energy comparison value.

The third step in the process of distinguishing the audio command from background noise is to compare the first energy content (of the command words) to the second energy content (of the background noise) and to arrive at an “energy comparison value.” The patent teaches that the energy comparison value may be expressed as either difference in energy content (e.g. $E_1 - E_2 = E_D$) or a ratio between the two energy contents (e.g. E_1/E_2) “or some other means by of comparing the energies.” *See* '630 patent, col. 12, l. 1-3.

ScanSoft's proposed construction of this term is off the mark. The term does not say anything about generating a signal or about a “characteristic value representative of the correspondence between the first and second energy contents.” It is not clear what these words

mean. There is no support for them in the written description. They should be rejected and Voice Signal's construction should be adopted.

Claim Term	ScanSoft's Proposed Construction	Voice Signal's Proposed Construction
[5] preventing the generation of said at least one control signal when said energy comparison value is below a predetermined level.	The control signal cannot be created if the energy comparison value is less than some preset value.	Voice Signal accepts ScanSoft's Proposed Construction.

Voice Signal agrees with ScanSoft's construction of this last element (although the parties disagree on the definitions of the component parts). The point of the analysis in claim 14 is to compare the energy of the command word portion(s) of an audio command with the energy of the background noise. The last step in the process is to prevent the generation of a control signal and thus to prevent the activation of an electrical device if the energy comparison value (as defined above) is below a predefined level. Stated differently, if this comparison does not sufficiently distinguish the command word(s) from the background noise, the system treats the entire received data as background noise (i.e. not as an audio command) and prevents the generation of a control signal. The written description states

if the sound data does not contain enough energy to meet the established energy difference [between commands and background noise] then the micro-controller assumes that the whole sound data is background noise and does not trigger the system.

'630 patent, col. 11, l. 64 – col. 12, l. 1.

Claim 16

Claim 16 reads, in its entirety, as follows:

A method of activating an electrical device through at least one audio command from a user, the method comprising the steps of:

- [1] recording speech recognition data having a command word portion and a pause portion, each of the speech recognition data portions being at least one syllable in length;
- [2] receiving at least one audio command from a user, the at least one audio command having first and second command word portions and a first, second and third pause portions, each of the audio command portions being at least one syllable in length, said second pause portion having one syllable in duration before said first command word portion and said third pause portion having one syllable in duration after said second command word portion;
- [3] comparing said command word portion and said pause portion of said at least one received audio command with said command word portion and said pause portion, respectively, of said speech recognition data;
- [4] generating at least one control signal based on said comparison;
- [5] controlling power delivered to an electrical device in response to said at least one control signal for operating the electrical device in response to said at least one received audio command.

Claim Term	ScanSoft's Proposed Construction	Voice Signal's Proposed Construction
A method of activating an electrical device through at least one audio command from a user, the method comprising the steps of:	<p>A process for turning on an electrical device</p> <p>An audio command is a word, phrase, or numeric digit, along or in combination with other words, phrases, and/or digits which is spoken by a user to turn on the electrical device</p>	<p>“activating” means “to make active or more active”</p> <p>“audio command” means an audible combination of command words and pauses that activate an electrical device</p> <p>A process for making an electrical device active or more active in response to at least one audio command</p>

Voice Signal's arguments in support of its construction of the claim terms “activating” and “audio command” are set forth at page 7 above.

Claim Term	ScanSoft's Proposed Construction	Voice Signal's Proposed Construction
[1] recording speech recognition data having a command word portion and a pause portion, each of the speech recognition data portions being at least one syllable in length;	<p>The "speech recognition data" is the speech model information that unknown input speech is compared to.</p> <p>A pause refers to a momentary stop in an ongoing event. The patent and its prosecution history define a syllable as a length of time at least about 200 milliseconds long. A "pause portion" is that part of the speech recognition data comprising a pause.</p> <p>Thus, this claim element requires that the speech model information includes a momentary stop in speech that is at least about 200 milliseconds long.</p>	<p>"speech recognition data" is the speech model data to which the received audio command is compared.</p> <p>the "pause portion" of the speech recognition data is the speech recognition data, or template, for the portion of the audio command that is not the command word portion.</p> <p>"at least one syllable in length" means "within the duration range for one syllable of speech."</p>

Voice Signal's arguments in support of its construction of the claim terms "speech recognition data" "pause portion" and "at least one syllable in length" are set forth at pages 8-11 above.

Claim Term	ScanSoft's Proposed Construction	Voice Signal's Proposed Construction
[2] receiving at least one audio command from a user, the at least one audio command having first and second command word portions and a first, second and third pause portions, each of the audio command portions being at least one syllable in length, said second pause portion having one syllable in duration before said first command word portion and said third pause portion having one	<p>The system gets from a user an audio command that has five different parts, all of which are at least 200 milliseconds long. Two word parts are separated by one syllable part. And two other different syllable parts respectively precede and follow the word parts.</p>	<p>"audio command" means an audible combination of command words and pauses that activate an electrical device.</p> <p>"at least one syllable in length" means "within the duration range for one syllable of speech."</p> <p>The audio command received from the user has at least two command word portions and</p>

syllable in duration after said second command word portion;		at least three pause portions, the pause portions of which are in between (1 st pause portion), before (2 nd pause portion) and after (3 rd pause portion) the command words.
--	--	--

Voice Signal's arguments in support of its proposed construction for the claim terms "audio command" and "at least one syllable in length" are set forth at pages 7, 11 above.

The method of claim 16 involves the receipt of an audio command that contains at least five parts. It includes two command word portions and three pause portions, as follows:⁵

<pause 2> command <pause 1> command 2 <pause 3>

As the written description explains, by utilizing the pause portions before and after the command word portions (pause 2 and pause 3), in addition to the command word portions and the pause portion in-between the command word portions, the system increases the number of syllables, and, therefore, the amount of data, being used to match audio commands to speech recognition data (the prerecorded templates). This improves the accuracy, or performance measure (called "Figure of Merit" or "FOM" in the written description, col. 1, l. 47-48)⁶ of the method.

ScanSoft's proposed construction agrees that in claim 16 the method employs audio commands consisting of five parts. However, it defines those parts as "word parts" and "syllable parts." This is not only wrong, but unnecessary. The claim calls the components "pause portions" and "command word portions." There is no purpose served by introducing new, undefined terms to the claim.

⁵ Compare claim 13 (which depends from claim 7 and adds a second pause portion before a first command word portion and a third pause portion after a second command word).

⁶ The Figure of Merit (FOM) is the average keyword detection over the range of 1-10 false alarms per keyword per hour. '630 patent, col. 1, l. 47-48.

Claim Term	ScanSoft's Proposed Construction	Voice Signal's Proposed Construction
[3] comparing said command word portion and said pause portion of said at least one received audio command with said command word portion and said pause portion, respectively, of said speech recognition data;	This claim element is indefinite and fails to satisfy the requirements of 35 U.S.C. §112. Therefore, it cannot be correctly construed.	Comparing the command word portions and the pause portions of the received audio command to the command word portions and the pause portions of the speech recognition data.

ScanSoft states that this third claim limitation of claim 16 is indefinite and therefore fails to satisfy 35 U.S.C. §112. It does not state the basis for its contention.

35 U.S.C. §112 ¶ 2 requires:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

When an applicant fails to satisfy Section 112 ¶ 2, the claim is said to be “indefinite.” The determination of whether a claim is indefinite is to be made by the Court. The question for the Court is whether a person skilled in the art would understand the bounds of the claim when read in light of the specification. *Exxon Research and Engineering Co. v. U.S.*, 265 F.3d 1371, 1375 (Fed. Cir. 2001) *Credle v. Bond*, 25 F. 3d 1566E, 1577 (Fed. Cir. 1997), *Miley Lab, Inc. v. Shandan, Inc.*, 997 F. 2d 870, 875 (Fed. Cir. 1993). Where the claim is amenable to interpretation by the Court in light of the specification, it should be not be found indefinite. *Id.*

The specification of the '630 patent renders clear anything which may be indefinite about this limitation in claim 16. As explained above, claim 16 describes a method using audio commands that include three pause portions and two command word portions. This element of claim 16 refers to “said command word portion” and “said pause portion.” The antecedent limitation describes two command word portions and three pause portions. Presumably, ScanSoft's indefiniteness argument is based on this inconsistent reference, which is quite

evidently a mistake. It is evident from the words of the claim itself and from the specification that the limitation at issue should have used the plural term portions in each instance rather than the singular term portion.

To the extent that there is any ambiguity created by the use of the singular (portion) rather than the plural (portions) in this limitation of claim 16, it is resolved in light of the specification, which describes comparing and matching command words (plural) and pauses (plural) in connection with the five part method. '630 patent, col. 10, l. 56-60. There can be no question that one of ordinary skill in the art would understand this limitation to be interpreted in the plural, not the singular, thereby meeting the requirements of § 112 ¶ 2.

“When a harmless error in a patent is not subject to reasonable debate, it can be corrected by the court.” *Hoffer v. Microsoft Corp.*, 2005 WL927148*4 (Fed. Cir. Apr. 22, 2005) (citing *Novo Indus., L.P. v. Micro Molds Corp.*, 350 F.3d 1348, 1356-57 (Fed. Cir. 2003)). Here, it is clear that the word “portion” in claim 16 should be the plural “portions.” That is how it should be construed.

Claim Term	ScanSoft's Proposed Construction	Voice Signal's Proposed Construction
[4] generating at least one control signal based on said comparison; and	A control signal is a signal that directs or manages the operation of the electrical device that is controlled by the control signal.	Voice Signal accepts ScanSoft's proposed construction.

Claim Term	ScanSoft's Proposed Construction	Voice Signal's Proposed Construction
[5] controlling power delivered to an electrical device in response to said at least one control signal for operating the electrical device in response to said at least one received audio command.	Electrical power is turned on or off, or increased or decreased, to an electrical device in response to the control signal generated in response to a five part input audio command of the type required above.	Electrical power is turned on or off, or increased or decreased, to an electrical device in response to the control signal generated in response to the received audio command.

The parties are in substantial agreement on the interpretation of this element. The only substantive difference is ScanSoft's insertion of the words "five part input audio command of the type described above." Since audio command is a term to be construed by the Court, there is no need for this added verbiage.

CONCLUSION

For the reasons set forth herein, Voice Signal's proposed claim constructions should be adopted.

Respectfully submitted,

LAURENCE S. GILLICK, ROBERT S. ROTH,
JONATHAN P. YAMRON, MANFRED G.
GRABHERR and VOICE SIGNAL
TECHNOLOGIES, INC.

By their attorneys,

/s/ Wendy S. Plotkin

Robert S. Frank, Jr. (BBO No. 177240)
Sarah Chapin Columbia (BBO No. 550155)
Paul D. Popeo (BBO No. 567727)
Paul E. Bonanno (BBO No. 646838)
CHOATE, HALL & STEWART LLP
Exchange Place
53 State Street
Boston, MA 02109
(617) 248-5000

Dated: May 6, 2005